Embracing the Imaginary:
Deep Complex-valued Networks for Heart Murmur Detection

Erika Bondareva, Georgios Rizos, Jing Han, Cecilia Mascolo

University of Cambridge, Cambridge, UK

Abstract

Machine learning for automated heart auscultation offers a scalable solution with the potential to broaden the accessibility of vital healthcare services. While conventional short-time Fourier transform-based audio representations contain both amplitude and phase information, the vast majority of proposed machine audition methods only consider the magnitude, discarding the phase information. In this work, we explore, for the first time, the potential of complex-valued neural networks (CVNNs) for heart sound classification, leveraging all available input information to derive complex representations from sound segments.

We showcase the effectiveness of complex-valued neural networks for sound analysis by directly comparing them with real-valued counterparts of our employed neural architectures. On the patient-independent test set of the PhysioNet 2022 Challenge dataset, a complex-valued treatment of two neural network architectures — including HMS-Net, the winning model of PhysioNet 2022 — leads to a consistent 1% absolute improvement in murmur detection accuracy compared to the real-valued baseline. Ensemble model using the non-complex and complex varieties of the two network architectures exhibits a further 6% improvement in sensitivity of murmur and 15% increase in sensitivity of unknown. This highlights the benefits of leveraging the complex domain in deep learning for heart sound analysis.

1. Introduction

Heart auscultation, while cost-effective and broadly accessible, has limited accuracy due to its reliance on human hearing. As a result, in clinical practice newer diagnostic tools that require less training are gaining prominence for their comparable or superior precision. To retain the accessibility of auscultation while eliminating the need for extensive auditory training, the research community has been increasingly exploring the application of machine learning techniques for audio-based diagnostics.

The field of automated cardiac auscultation has seen a wide range of approaches, reflecting broader trends in the machine learning domain. Initially, researchers focused on signal processing methods, such as signal envelopes and feature-engineering approaches [1]. Traditional machine learning models and probabilistic models (such as hidden Markov models) emerged as the field matured [2, 3]. More recently, there has been a marked shift towards deep learning-based methods [4], specifically focusing on improving heart murmur detection accuracy.

The PhysioNet 2022 Challenge significantly contributed to this area by releasing the largest heart sound dataset to date [5]. This challenge invited the design of algorithms for heart sound classification, focusing specifically on murmur detection and outcome prediction [6]. In our work, we concentrate on the murmur detection task, as early identification of murmurs can be crucial for timely intervention and effective management of potential heart conditions.

Humans perceive audio in the time domain. However, the convention for signal processing and machine learning is to convert audio into the frequency domain. Since the Discrete Fourier transform (DFT) is subject to Heisenberg's uncertainty principle, the Short-time Fourier transform (STFT) is frequently used to overcome this problem. STFTs, as the name suggests, capture frequency amplitudes over short windows, thus creating a two-dimensional matrix. Since DFT is intrinsically complex, STFT inherits this property. Spectrograms, derived from STFT by calculating the magnitude, are widely used due to their compatibility with vision-based neural architectures.

Some of the recent research efforts focused on utilising spectrograms alongside deep learning to provide the best performance seen to date [7, 8]. The winning entry of the PhysioNet 2022 Challenge, HMS-Net, utilised real-valued multiscale spectrograms in a hierarchical convolutional network for effective murmur detection and identification of poor-quality samples (i.e. unknown class) [9].

However, real and imaginary components of the DFT of an audio signal are statistically dependent on each other, which is not captured upon extracting a magnitude spectrogram from STFT. By using a model that is able to directly process raw, complex-valued STFTs, we may be able to harness all the available information. In a complex space, neurons process data in two dimensions, which allows the network to learn more nuanced relationships in the data. In addition, the added constraints of this approach may yield a more consistent and stable performance.

Complex-valued neural networks (CVNNs) have a long history [10], but they are not commonly used. In acoustics, they have been explored for audio denoising [11] and music transcription [12], but not for automated auscultation.

For the first time, this work explores the potential of CVNNs for murmur detection. Specifically, the contributions of this paper are as follows:

- We implement and replicate a real-valued HMS-Net [9], and introduce its complex-valued variant.
- We demonstrate the effectiveness of CVNNs in heart sound analysis by examining two distinct architectures — a convolutional neural network (CNN) and the HMS-Net. Across both, our results consistently indicate a 1% improvement in accuracy when transitioning from their real-valued variants to their complex-valued counterparts.
- We achieve a decrease in standard deviation for five-fold cross-validation, demonstrating a more stable performance of complex models in comparison to their real-valued counterparts across different folds.
- An ensembling approach using the non-complex and complex varieties of the CNN and HMS-Net models shows a 6% improvement in the sensitivity of murmur and a 15% increase in the sensitivity of unknown.

2. Methods

2.1. Complex-valued Neural Networks

In this work, we explore a fully complex-valued neural network (CVNN) approach. If $z, z' \in \mathbb{C}$ are the input and output of a complex-valued layer, respectively, we could compute the output z' as: $z' = Wz + b$, where W and b are the complex-valued weights and biases, respectively. This way the complex network variants mirror their real-valued counterparts, with individual elements (layers, dropout, pooling, activations, and backpropagation) being replaced by their complex equivalents using the NEGU93/CVNN Python library [13].

In CVNNs, constraints arise from the coupling of real and imaginary parts, leading to interdependent activations in the complex neurons. Holomorphic activation functions impose additional mathematical constraints, contributing to stable learning dynamics [14]. Despite the increase in parameters due to complex numbers, this coupling can effectively reduce the degrees of freedom, constituting a form of regularisation. These constraints can make CVNNs more robust and less prone to overfitting, potentially resulting in more stable and consistent performance.

We evaluated multiple complex activation functions, selecting the best-performing ones for our final results. The intermediate layers utilise cartReLU activation, in which a rectified linear unit is applied to both real and imaginary parts [12]. For the output layers, we applied softmax to the real and imaginary parts separately before averaging.

2.2. Deep Network Architectures

In order to facilitate a fair comparison of CVNNs and their real-valued counterparts, we investigate two real-valued models: a vanilla convolutional neural network (**CNN**) with six layers, and a hierarchical multi-scale convolutional network, **HMS-Net**, proposed by one of the winners of the PhysioNet Challenge 2022 [9].

The architecture of the **CNN** model consists of an input layer, followed by six convolutional layers, each increasing in filter depth, each with ReLU activation function. Each convolutional layer is succeeded by a 15% dropout and max-pooling for dimensionality reduction. Finally, there is a flattening layer to transition from convolutional segments to a dense layer. The final output layer has a softmax activation. The Adam optimisation algorithm with a learning rate of 0.0001 is used for training.

HMS-Net is a variant of ResNet [15], in that it processes spectrograms across varied scales. There is a module dedicated to learning latent representations for each of these inputs. After appropriate sub-sampling, these representations are progressively merged via concatenation, before a joint processing step, followed by global average pooling of the time and frequency dimensions.

2.3. Dataset and Features

For this study, we used the PhysioNet 2022 publicly released training dataset [5] which contains heart sound (HS) labels for each patient, as well as for individual samples. The dataset comprises 695 patients without murmurs, 179 patients with a murmur present, and 68 unknown.

We adopt a preprocessing methodology similar to the winning entry of the PhysioNet 2022 Challenge [9]. Specifically, we filtered and downsampled the audio to 2000 Hz. All the recordings were segmented into 3 s overlapping windows with a hop length of 1 s. Each segment was treated as a separate sample for the model training.

For feature extraction, we computed STFT at three different scales, denoted as x1, x0.5, and x0.25. These scales had varying DFT bins, window lengths, and hop lengths. The x1 scale used 446 bins with window and hop lengths of 200 and 27 samples, respectively; the x0.5 scale utilised 222 bins with 100 and 54 samples; and the x0.25 scale had 110 bins with window and hop lengths set to 50 and 108 samples. While the HMS-Net employed all three scales, the basic neural network only used the x1 scale.

We further use an audio quality metric [9] (frequency energy ratio between 20-200 Hz and 0-1000 Hz) to retain all murmur samples and re-label poor quality normal samples (i.e. with a power spectral density ratio below 0.3) as unknown. Moreover, to ensure that the model is exposed to a relatively balanced dataset, we upsample the murmur class in the training set by a factor of three.

For real-valued architectures, we derived the magnitude from the STFTs using $\mathit{Magnitude} = \sqrt{\Re^2 + \Im^2}$. For the complex-valued models, however, raw STFT was utilised.

Since murmur sensitivity is one of the most clinically important performance metrics, we focused on boosting it by implementing an ensemble algorithm which combined patient predictions of all four models. It produced the final result based on two rules: if at least one model predicts the patient to have a murmur, the final result is murmur; and the same rule was then applied for unknown.

2.4. Prediction and Evaluation

The prediction and prediction aggregation process for both models, in their real and complex-valued forms, was conducted in a tiered manner:

- **Segment-level predictions**: For each 3 s segment, initial predictions were generated using the respective models.
- **Audio recording-level aggregation**: This represents the diagnosis for one auscultatory location per patient, determined by selecting the most frequent prediction among all segments of a recording.
- **Patient-level aggregation**: Based on the aggregated results of audio recordings, a patient is diagnosed with a murmur if at least one location indicates its presence. If half or more recordings for a patient are classified as unknown, the overall diagnosis defaults to unknown.

The evaluation was conducted using patient-independent five-fold cross-validation, adopting an 80:20 training-to-testing ratio. Results are presented as the mean and standard deviation across the five folds. We reported precision and sensitivity for each class, the accuracy of known (excluding the unknown class) and the total accuracy.

3. Results and Discussion

In order to confirm our hypothesis — that using STFTs directly within a complex-valued neural network architecture allows for a richer understanding of amplitude-phase relationships — we explore the performance of four distinct architectures: non-complex and complex CNN, and non-complex and complex HMS-Net. The results are summarised in Table 1.

For CNN models, we see that the complex model outperforms its non-complex counterpart across all metrics, except the sensitivity of unknown. It is important to note that all unknown samples consist of normal or murmur heart sounds with elevated levels of noise. Therefore, the dip in unknown sensitivity when using the complex model might be attributed to the following reasons: the complex model is more robust to noise, thereby efficiently sorting the noisy samples into normal or murmur classes, or the complex model is more sensitive towards class imbalance. Since unknown sounds is such a minority class, we can reasonably expect the performance to fluctuate a lot. Therefore, going beyond the evaluation scheme of the PhysioNet 2022 challenge, we also report the accuracy of the known, which is between the two bigger classes. Both accuracy of known and total accuracy are higher for the complex variant than for the real-valued counterpart.

HMS-Net was a tied winner in the murmur detection task during the PhysioNet 2022 Challenge. In our effort to faithfully replicate this pipeline, our closest attempt achieved an average accuracy of 82%. This discrepancy with the reported average accuracy of 83.7% may stem from variations in the train-test split or minor differences in final processing and model training. However, when comparing our leading HMS-Net variant with its complex-valued counterpart, we demonstrated an equivalent improvement of 1% in favour of the complex variant.

Overall, the HMS-Net outperformed the CNN model in terms of sensitivity for both murmur and unknown classes. This could be attributed to the HMS-Net's multi-scale processing capability, which enables the model to capture both granular and broader audio features. The enhanced total accuracy of the complex models could be explained by the ability of complex-valued networks to learn from all available information encoded in raw STFTs, resulting in more comprehensive representations of the heart sounds.

It is also worth noting that the complex variants exhibited lower standard deviations compared to their non-complex counterparts. This observation aligns with the hypothesis that the intrinsic constraints of the complex model contribute to a more stable performance.

The ensemble model achieves the highest sensitivity in murmur and unknown, at 74% and 62%, respectively. Since the total accuracy has remained at 83%, we argue that the ensemble model provides significant improvement due to the clinical importance of the murmur sensitivity.

4. Conclusions

We have explored complex-valued neural networks for heart murmur detection, directly leveraging STFTs. Our results show that the complex-valued approach, especially when adapting the HMS-Net architecture, outperforms its real-valued counterparts across most metrics.

Our research suggests the potential benefits of a methodological shift: complex-valued neural networks might improve the performance of an existing real-valued network.

The improved sensitivity for both murmur and unknown classes using the ensemble model indicates that real and complex-valued architectures correctly identify distinct

Table 1. Final results for the vanilla CNN model, HMS-Net, and the Ensemble model for both real and complex-valued inputs and models. The results are reported for a 5-fold cross-validation as mean $\pm$ stdev.

	CNN		HMS-Net		Ensemble model
	non-complex	complex	non-complex	complex	
Precision of normal	0.87 ± 0.02	0.87 ± 0.02	0.90 ± 0.03	0.89 ± 0.03	0.92 ± 0.04
Precision of murmur	0.93 ± 0.09	0.95 ± 0.08	0.88 ± 0.07	0.92 ± 0.06	0.86 ± 0.07
Precision of unknown	0.31 ± 0.16	0.33 ± 0.12	0.29 ± 0.18	0.30 ± 0.15	0.34 ± 0.13
Sensitivity of normal	0.91 ± 0.02	0.93 ± 0.03	0.90 ± 0.03	0.92 ± 0.02	0.87 ± 0.03
Sensitivity of murmur	0.60 ± 0.11	0.63 ± 0.10	0.68 ± 0.17	0.64 ± 0.11	0.74 ± 0.12
Sensitivity of unknown	0.44 ± 0.23	0.40 ± 0.19	0.47 ± 0.30	0.46 ± 0.31	0.62 ± 0.22
Accuracy of known	0.85 ± 0.03	0.87 ± 0.03	0.85 ± 0.05	0.86 ± 0.03	0.84 ± 0.04
Total accuracy	0.82 ± 0.03	0.83 ± 0.02	0.82 ± 0.04	0.83 ± 0.02	0.83 ± 0.05

sets of murmur samples. Since these models focus on disparate aspects of the signal for their final predictions, ensembles combining real and complex-valued architectures offer a promising avenue for performance improvement.

A promising area for future research is to compare architectures successful in other acoustic applications with their complex-valued versions, specifically for murmur detection, as well as for other audio-based diagnostics.

Acknowledgments

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L015889/1 for the Centre for Doctoral Training in Sensor Technologies and Applications, and by ERC Project 833296 (EAR).

References

[1] Sun S, Jiang Z, Wang H, Fang Y. Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified hilbert transform. Computer Methods and Programs in Biomedicine 2014;114(3):219–230.

[2] Springer DB, Tarassenko L, Clifford GD. Logistic regression-hsmm-based heart sound segmentation. IEEE Transactions on Biomedical Engineering 2016;63(4):822–832.

[3] Whitaker BM, Suresha PB, Liu C, Clifford GD, Anderson DV. Combining sparse coding and time-domain features for heart sound classification. Physiological Measurement 2017;38(8):1701.

[4] Bizopoulos P, Koutsouris D. Deep learning in cardiology. IEEE Reviews in Biomedical Engineering 2019;12:168–193.

[5] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C, Jorge A, Mattos S, Hatem T, Tavares T, Elola A, Rad AB, Sameni R, Clifford GD, Coimbra MT. The CirCor DigiScope dataset: from murmur detection to murmur classification. IEEE Journal of Biomedical and Health Informatics 2021;26(6):2524–2535.

[6] Reyna MA, Kiarashi Y, Elola A, Oliveira J, Renna F, Gu A, Perez-Alday EA, Sadr N, Sharma A, Mattos S, Coimbra MT, Sameni R, Rad AB, Clifford GD. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. medRxiv 2022;.

[7] Lee J, Kang T, Kim N, Han S, Won H, Gong W, Kwak IY. Deep learning based heart murmur detection using frequency-time domain features of heartbeat sounds. In 2022 Computing in Cardiology (CinC), volume 498. 2022; 1–4.

[8] Lu H, Yip JB, Steigleder T, Grießhammer S, Heckel M, Jami NVSJ, Eskofier B, Ostgathe C, Koelpin A. A lightweight robust approach for automatic heart murmurs and clinical outcomes classification from phonocardiogram recordings. In 2022 Computing in Cardiology (CinC), volume 498. 2022; 1–4.

[9] Xu Y, Bao X, Lam HK, Kamavuako EN. Hierarchical multi-scale convolutional network for murmurs detection on pcg signals. In 2022 Computing in Cardiology (CinC), volume 498. 2022; 1–4.

[10] Hirose A. Applications of complex-valued neural networks to coherent optical computing using phase-sensitive detection scheme. Information Sciences Applications 1994; 2(2):103–117.

[11] Hirose A, Yoshida S. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. IEEE Transactions on Neural Networks and Learning Systems 2012;23(4):541–551.

[12] Trabelsi C, Bilaniuk O, Serdyuk D, Subramanian S, Santos JF, Mehri S, Rostamzadeh N, Bengio Y, Pal CJ. Deep complex networks. CoRR 2017;.

[13] Barrachina JA. Negu93/cvnn: Complex-valued neural networks, November 2022. URL `https://doi.org/10.5281/zenodo.7303587`.

[14] Sarroff AM. Complex neural networks for audio. Ph.D. thesis, Dartmouth College, Hanover, New Hampshire, May 2018.

[15] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 770–778.

Address for correspondence:

Erika Bondareva: eb729@cam.ac.uk
15 JJ Thomson Ave, Cambridge, UK, CB3 0FD